Ensuring and Improving Information Quality for Earth Science Data and Products: ESIP Information Quality Cluster

Hampapuram Ramapriyan[1] (Hampapuram.Ramapriyan@ssaihq.com)
David Moroni[2] (David.F.Moroni@jpl.nasa.gov)
Ge Peng[3] (Ge.Peng@noaa.gov)
[1]NASA Goddard Space Flight Center and Science Systems and Applications, Inc.
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
[3]NOAA's Cooperative Institute for Climate and Satellites - North Carolina (CICS-NC) and NOAA's National Centers for Environmental Information (NCEI)
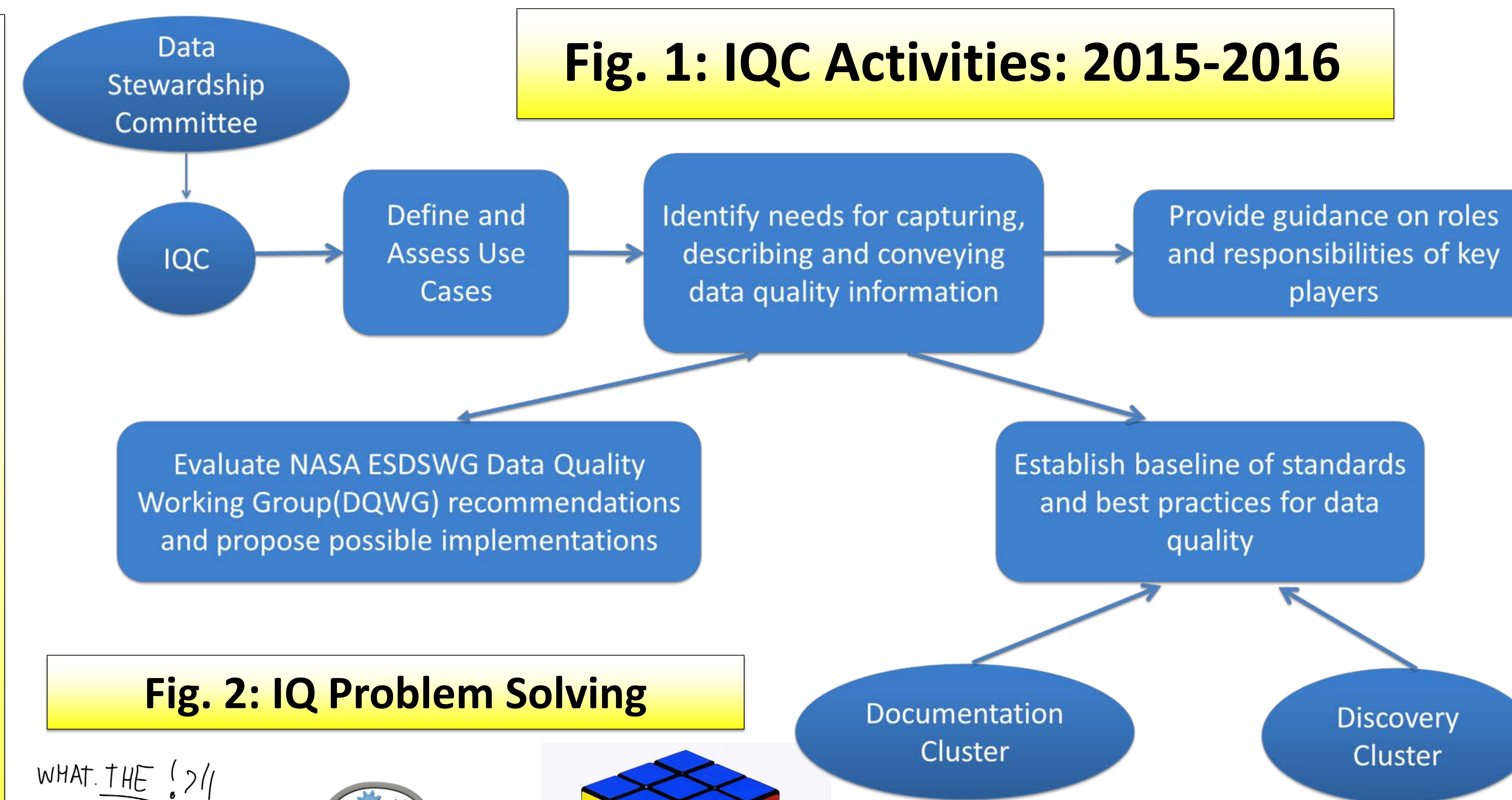
Abstract: Quality of Earth science data products is always of concern to users regardless of the type of products. Science quality, product quality, stewardship quality and service quality are four distinct aspects of information quality. With increasing requirements on ensuring and improving information quality coming from multiple government agencies and throughout industry, there have been considerable efforts toward improving information quality during the last decade. Given this rich background of prior work, the Information Quality Cluster (IQC), has been active with membership from multiple government agencies, institutions, and organizations. The vision of IQC is **"to become internationally recognized as an authoritative and responsive resource of information and guidance to data providers on how best to implement data quality standards and best practices for their science data systems, datasets, and data/metadata dissemination services."** IQC's objectives and activities, aimed at ensuring and improving information quality for Earth science data and products, are discussed briefly. During 2016, the focus of IQC was on the development and assessment of use cases to identify issues related to collecting and conveying quality information to users, and recommending improvements for implementation by data producers and data distributors. During 2017, the emphasis of IQC has shifted towards increased collaborations with other ESIP clusters and entities outside the ESIP in the form of including invited speakers at IQC meetings and exchanging information.

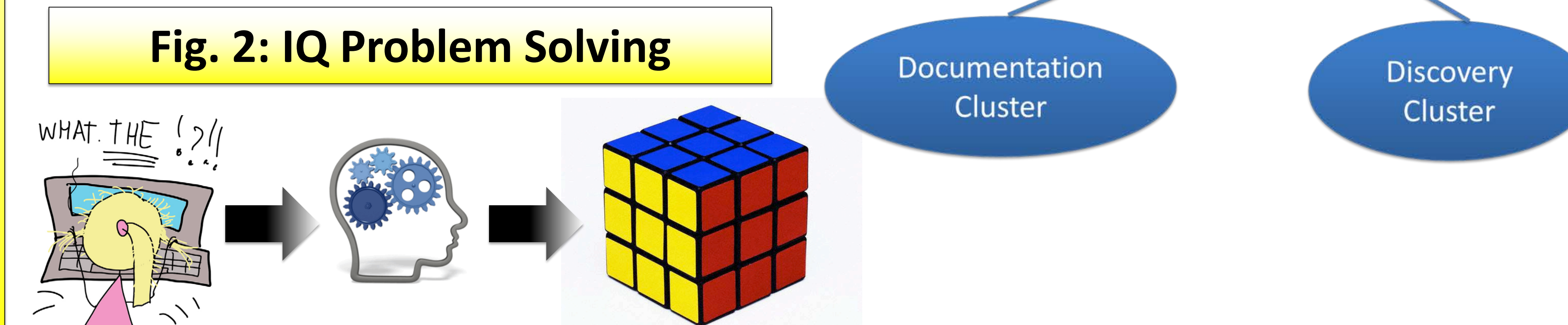## Fig. 1: IQC Activities: 2015-2016

Fig. 2: IQ Problem Solving

Information Quality Aspects:

1. **_Science_**: accuracy, precision, uncertainty, validity, and suitability for use (fitness for purpose).
2. **_Product_**: degree to which scientific quality is assessed and documented; how accurate, complete and up-to-date metadata and documentation are; manner in which data and metadata are formatted; degree to which associated information is published and traceable throughout the data lifecycle.
3. **_Stewardship_**: how well data are being managed and preserved.
4. **_Service_**: how easy it is for users to discover, get, understand, trust, and use a given data product along with its metadata, as well as ensuring that an archive has the requisite knowledge base and people functioning as subject matter experts available to help its data users.

Fig. 3: Scope of Mutual Influence and Sources of Domain Knowledge
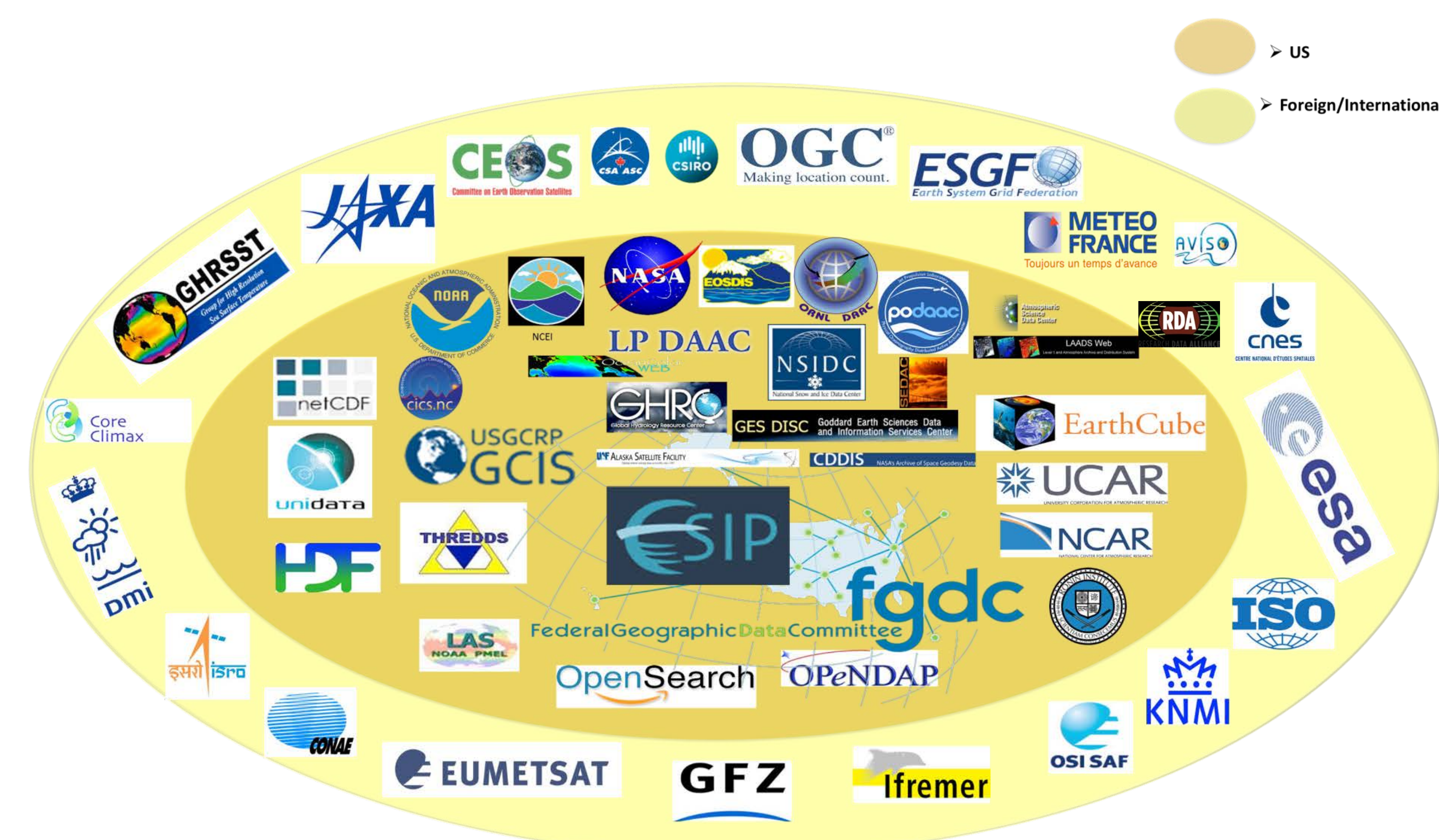
2016 IQ Use Cases Evaluation

1. **Dataset Rice Cooker Theory** – Bob Downs, David Moroni, and Joseph George
2. **Appropriate Amount/Extent of Documentation for Data Use** – Ge Peng, Steve Olding, and Lindsey Harriman
3. **Improving Use of SBC LTER Data Portal** – Margaret O'Brien, Sophie Hou, Ross Bagwell, and Hampapuram Ramapriyan.
4. **Citizen Science** – Ruth Duerr, Han Qin, Chung-Lin Shie, and Bhaskar Ramachandran
   See http://wiki.esipfed.org/index.php/File:Summary_of_Evaluations_of_Use_Cases.pdf for details.

2017 Activities

Winter ESIP Meeting – IQC session on Fostering Collaborations: Applications-Oriented Presentations*
- Jeff Campbell (Agriculture and Climate Cluster/USDA) – *Agricultural Research Perspectives*
- Karen Moe (Disaster Lifecycle Cluster/NASA GSFC) – *ESIP Disasters Lifecycle Cluster and Information Quality*
- Pierre Guillevic (U of MD) – *CEOS Land Product Validation Overview and Goals*
- Robert Ferraro (JPL) – *Obs4MIPS*

Monthly telecons – Invited Speakers*
- April – **Aaron Friesz** & Lindsey Harriman (LP DAAC) – *Demystifying MODIS quality data: Determining data usability via quality layers.*
- May – Helena Cousijn, Claire Austin & **Michael Diepenbroek (PANGAEA)** – *Assessment of Data Fitness for Use*
- June – **Nicholas Car (Geoscience Australia)** – *Data Reuse Fitness Assessment Using Provenance*
- July – **Mark Reese (Element 84)** – *Usability Considerations in Conveying Data Quality Information*

Publication and Presentations*:
- Moroni, D., H. Ramapriyan, and G. Peng, "A Platform to Provide International and Inter-Agency Support for Data and Information Quality Solutions and Best Practices", *International Ocean Vector Winds Science Team Meeting*, 2–4 May 2017, San Diego, CA, USA.
- Peng, G., Ramapriyan, H., and D. Moroni, "The State of Building a Consistent Framework for Curation and Presentation of Earth Science Data Quality", *The ESIP winter meeting*, 11–13 January 2017, Bethesda, MD, USA.
- Ramapriyan, H., G. Peng, D. Moroni, C-L Shie, "Ensuring and Improving Information Quality for Earth Science Data and Products", *D-Lib Magazine*, July 2017, DOI: https://doi.org/10.1045/july2017-ramapriyan

* Presentations at http://wiki.esipfed.org/index.php/IQ_Presentations